BINAURAL VS MONAURAL STI MEASUREMENTS

Constant Hak
Eindhoven University of Technology, the Netherlands
email: c.c.j.m.hak@tue.nl

Remy Wenmaekers
Level Acoustics & Vibration, Eindhoven, the Netherlands
email: remy.wenmaekers@levelav.nl

Han Vertegaal
Acoustics Engineering, Boxmeer, the Netherlands
email: h.vertegaal@acoustics-engineering.com

According to the IEC 60268-16 standard, the STI model for the prediction of speech intelligibility is based on monaural listening, measured using an omnidirectional microphone. However, the same standard mentions the subjective advantages of binaural listening to speech intelligibility. This study aims at quantifying such advantages, for example under various directive sound field conditions. To this end, the STI-values from binaural impulse responses using a stepwise rotating HATS (Head and Torso Simulator) in a concert hall, are compared to the STI using an omnidirectional microphone. The resulting polar diagrams clearly show a directivity in the binaural STI, and the best of the two STI-values from the HATS, referring to the so-called Best Ear method, indeed tends to exceed the monaural STI. Differences in STI of up to 0.1 are found, exceeding 3 times the Just Noticeable Difference for STI. Also, the angle-independent single-number quantities, obtained by averaging each two HATS channels, are found to exceed those from omnidirectional measurements.

Keywords: speech intelligibility, binaural STI, HATS rotation

1. Introduction

The IEC 60268-16 [1] standard describes how to predict the speech intelligibility by measuring the STI (Speech Transmission Index) [2] or its simplified derivative STIPA (STI for Public Address systems) [3]. The STI(PA) model is based on monaural listening and uses low frequency modulations ranging from 0.6 to 12.5 Hz in noise bands from 125 to 8 kHz as an average human 'speech rhythm'. Where the STI uses 14 modulation frequencies in each noise band, the STIPA only uses 2. The preservation of these modulations in the transmission path from speaker to listener are a measure of speech intelligibility. Modulation (or speech intelligibility) reduction is caused by background noise (SNR) and by the acoustic properties of the room, as modelled in equation (1) where m is the modulation index [-], f_m is the modulation frequency [Hz], T the ('straight line') reverberation time [s] and SNR the signal to noise ratio [dB].

$$m(f_m) = \frac{1}{\sqrt{1+\left(\frac{2\pi f_m T}{13.8}\right)^2}} \cdot \frac{1}{1+10^{-\frac{SNR}{10}}} \quad (1)$$

Although the STI model for the prediction of speech intelligibility is based on monaural listening, alternative measurements can be made with an artificial binaural ear/head simulator. The IEC standard mentions the subjective advantages of binaural listening to speech intelligibility. In their study, Wijngaarden and Drullman found an underestimation in predicting speech intelligibility when measuring the standard speech transmission index to predict speech intelligibility in binaural listening conditions [4]. Various directive sound field conditions caused by listener position, surface shape, sound reflection properties, directivity of PA source(s) and background noise elements may seriously impact a STI value. Significant improvement in STI evaluation is already obtained by simply doing a two-channel STI measurement using an artificial head and using the Best Ear method.

In this paper, the STI-values from binaural impulse responses measured in a concert hall using a 10 degree stepwise rotating HATS (Head and Torso Simulator) are compared to the STI-values using an omnidirectional microphone. The 'omni' and HATS values for all speech intelligibility parameters mentioned in the IEC 60268-16 standard are derived from binaural impulse responses obtained from a previous concert hall study [5]. The STI, STIPA and MTI(f), were determined for one omnidirectional sound source and two receiver positions, both with an omnidirectional microphone and a HATS.

In this case study (of a smooth acoustics concert hall) the differences between the omnidirectional microphone and HATS results caused by directivity effects of background noise (sources) can be neglected due to the high SNR. The differences in results are only caused by the hall itself and the directivity properties of the omni source for frequencies up from 1 kHz [6].

A description of the measurement setup and all results of this explorative/experimental investigation are presented in a set of polar plots and tables with angle-independent single-number quantities.

2. Measurements

2.1 Measurement conditions

To evaluate the difference between an omnidirectional microphone and a HATS on the measured room acoustic parameters, impulse response measurements were performed in the large concert hall of "Muziekgebouw Eindhoven" with a volume of approx. 14,400 m^3, an unoccupied stage floor and T_{empty} ≈ 2 s. Figure 1 gives an impression of the hall and a schematic floorplan with the source position S as indicated, placed on the major axis of the hall, and the microphone positions R1 and R2, where R1 is placed at approx. 5 m from the source, equal to the critical distance and R2 is placed at approx. 18 m (diffuse field). A complete description of the hall, including an impression and the acoustic properties, can be found in Internoise paper [5].

2.2 Measurement definitions and equipment

The *source direction* is defined in terms of the 'line of sight' of the HATS to the source, where the direction is specified as the angle between the line of sight and the major axis of the hall. For R1 this angle ≈ 32° while for R2 the angle ≈ 12°. The *stage viewing angle* is the angle between the line of sight of the HATS when pointed into the direction of the leftmost side of the stage, and the line of sight when pointed to the rightmost side of the stage. For R1 this angle ≈ 156° while for R2 the angle ≈ 60°. See figure 1 for clarification.

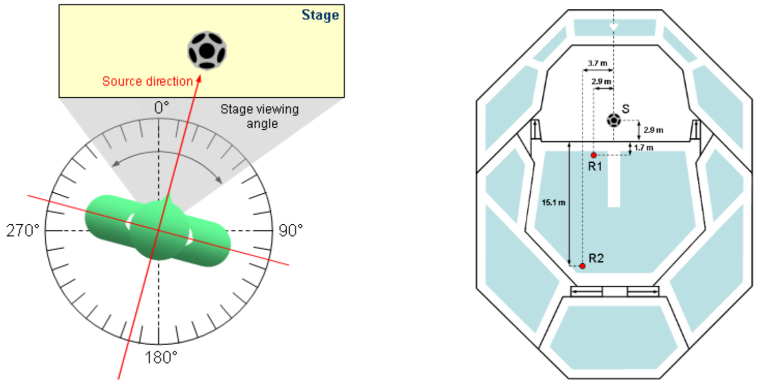


Figure 1: During the measurements the HATS from Brüel & Kjær turned in steps of 10°

2.3 Measurement equipment

The used measurement equipment can be found in Internoise paper [5]. Using a turntable, the HATS could rotate 360° around its vertical axis. After removal of the HATS, the omnidirectional microphone was placed at a position virtually centred between the ears of the HATS. For the STI-related parameters the indirect method was used obtained from IRs via a 5.37 s exponentially swept sine [7][8]. The decay ranges INR [9] for all impulse responses measured with this setup had an average of 60 dB for all frequency bands, with a minimum exceeding 45 dB. For all files the minimum SNR (Signal to Noise Ratio) was 30 dB for the 125 and 8000 Hz octave bands and 40 dB for 500, 1000 and 2000 Hz.

3. Results and discussion

For a single source position (S) and two microphone positions (R1 and R2), the differences between 'omnidirectional parameter values' and the same parameters measured with a HATS were investigated. Polar plots containing the measurement results can be found in figures 3 and 4, where 3 frequency ranges are distinguished: 'Low' as average over the 125 Hz and 250 Hz octave bands, 'Mid' as average over the octave bands 500 Hz, 1 kHz and 2kHz, and 'High' being the average over the 4 kHz and 8 kHz octave band values. For the speech intelligibility parameters MTI, STI and STIPA, described in IEC 60268-16 and (for completeness) the speech parameter D_{50}, described in ISO 3382-1 [10], the polar plots are given for both the omnidirectional microphone and the HATS at the same listener position. Here the HATS was rotated around its vertical axis in steps of 10 degrees, and at each step a measurement was performed. In addition to the polar plots, average numeric values were calculated and are presented in tables 6 through 9. Tables 6 and 7 contain the results for receiver position R1; tables 8 and 9 for position R2. Tables 6 and 8 list the differences between the omnidirectional microphone and the HATS. Tables 7 and 9 contain the absolute maximum differences between both ears of the HATS. Results are listed, averaged over a full 360° rotation of the HATS, averaged over a partial rotation over the stage viewing angle, and for the fixed source direction. Notable in the polar plots, both for position R1 (table 4: 'critical distance') and position R2 (table 5: 'diffuse field'), is the clear directivity for all speech parameters that were investigated. For both receiver positions (R1 and R2) Most parameter value differences between omnidirectional and HATS (left ear, right ear and averaged ear) are less than 1 JND [11]. The STI and the STIPA show more considerable average differences (Best Ear - Omnidirectional) of approx. 1.5 times the JND at both measurement positions up to max 3.4 times the JND at position R1 and 1.9 times the JND at position R2. Table 3 lists the JND values for the parameters used.

Table 3: JND values

EDT	T_{30}	D_{50}	MTI	STI(PA)
5%	10 %	0.05	0.03*	0.03

*Assumed to be equal to STI(PA) JND

Table 6: Parameter value differences between omnidirectional microphone and HATS for Receiver position **R1**, with sound-receiver-distance of 5 m and source direction of 32°

Receiver (Microphone)	Average over 360°									
	MTI (oct band Hz) 125	250	500	1k	2k	4k	8k	*STI*	*STIPA*	D_{50}*
Omnidirectional	0.538	0.540	0.575	0.551	0.583	0.671	0.588	0.586	0.581	0.541
Left ear HATS	0.539	0.572	0.599	0.531	0.601	0.664	0.596	0.593	0.587	0.537
Right ear HATS	0.539	0.572	0.599	0.534	0.603	0.670	0.589	0.594	0.587	0.541
Avg (L,R) HATS	0.539	0.572	0.599	0.532	0.602	0.667	0.593	0.594	0.587	0.539
Receiver (Microphone)	**Average over stage viewing angle**									
	MTI (oct band Hz) 125	250	500	1k	2k	4k	8k	*STI*	*STIPA*	D_{50}*
Omnidirectional	0.538	0.540	0.575	0.551	0.583	0.671	0.588	0.586	0.581	0.541
Left ear HATS	0.534	0.567	0.591	0.505	0.573	0.653	0.584	0.575	0.565	0.524
Right ear HATS	0.544	0.577	0.607	0.561	0.656	0.758	0.650	0.637	0.634	0.555
Avg (L,R) HATS	0.539	0.572	0.599	0.533	0.615	0.706	0.617	0.606	0.600	0.540
Receiver (Microphone)	**Sound source direction**									
	MTI (oct band Hz) 125	250	500	1k	2k	4k	8k	*STI*	*STIPA*	D_{50}*
Omnidirectional	0.538	0.540	0.575	0.551	0.583	0.671	0.588	0.586	0.581	0.541
Left ear HATS	0.530	0.569	0.571	0.509	0.613	0.737	0.596	0.598	0.588	0.515
Right ear HATS	0.540	0.556	0.591	0.559	0.662	0.758	0.598	0.628	0.626	0.538
Avg (L,R) HATS	0.535	0.562	0.581	0.534	0.638	0.748	0.597	0.613	0.607	0.527

Table 7: Maximum parameter value differences between the Left and the Right ear of a HATS for receiver position **R1**, with sound-receiver-distance of 5 m and source viewing angle of 32°

Viewing direction	$\lvert \text{Left ear} - \text{Right ear} \rvert_{max}$									
	MTI (oct band Hz) 125	250	500	1k	2k	4k	8k	*STI*	*STIPA*	D_{50}*
Avg 360°	0.000	0.000	0.000	0.003	0.002	0.006	0.007	0.001	0.000	0.004
Avg Stage viewing angle	0.010	0.010	0.016	0.056	0.083	0.105	0.066	0.062	0.069	0.031
Source direction	0.010	0.013	0.020	0.050	0.049	0.021	0.002	0.030	0.038	0.023

Table 8: Parameter value differences between omnidirectional microphone and HATS for Receiver position **R2**, with sound-receiver-distance of 18 m and source direction of 12°

Receiver (Microphone)	Average over 360°									
	MTI (oct band Hz) 125	250	500	1k	2k	4k	8k	*STI*	*STIPA*	D_{50}*
Omnidirectional	0.437	0.455	0.477	0.462	0.423	0.489	0.565	0.469	0.474	0.258
Left ear HATS	0.444	0.459	0.451	0.483	0.457	0.514	0.593	0.487	0.498	0.278
Right ear HATS	0.444	0.460	0.452	0.485	0.455	0.523	0.592	0.488	0.501	0.279
Avg (L,R) HATS	0.444	0.459	0.452	0.484	0.456	0.518	0.592	0.488	0.500	0.279
Receiver (Microphone)	Average over stage viewing angle									
	MTI (oct band Hz) 125	250	500	1k	2k	4k	8k	*STI*	*STIPA*	D_{50}*
Omnidirectional	0.437	0.455	0.477	0.462	0.423	0.489	0.565	0.469	0.474	0.358
Left ear HATS	0.447	0.439	0.442	0.509	0.473	0.541	0.655	0.506	0.520	0.290
Right ear HATS	0.442	0.464	0.452	0.465	0.456	0.492	0.531	0.472	0.489	0.268
Avg (L,R) HATS	0.445	0.451	0.447	0.487	0.465	0.516	0.593	0.489	0.505	0.279
Receiver (Microphone)	Sound source direction									
	MTI (oct band Hz) 125	250	500	1k	2k	4k	8k	*STI*	*STIPA*	D_{50}*
Omnidirectional	0.437	0.455	0.477	0.462	0.423	0.489	0.565	0.469	0.474	0.358
Left ear HATS	0.448	0.437	0.438	0.515	0.474	0.539	0.667	0.508	0.523	0.287
Right ear HATS	0.443	0.464	0.451	0.456	0.460	0.490	0.516	0.469	0.482	0.266
Avg (L,R) HATS	0.445	0.450	0.444	0.485	0.467	0.515	0.591	0.489	0.503	0.277

Table 9: Maximum parameter value differences between the Left and the Right ear of a HATS for receiver position **R2**, with sound-receiver-distance of 18 m and source viewing angle of 12°

Viewing direction	$\lvert \text{Left ear} - \text{Right ear} \rvert_{max}$									
	MTI (oct band Hz) 125	250	500	1k	2k	4k	8k	*STI*	*STIPA*	D_{50}*
Avg 360°	0.000	0.001	0.001	0.002	0.002	0.009	0.001	0.001	0.003	0.001
Avg Stage viewing angle	0.005	0.025	0.010	0.044	0.017	0.049	0.124	0.034	0.031	0.022
Source direction	0.005	0.027	0.013	0.059	0.014	0.049	0.151	0.039	0.041	0.021

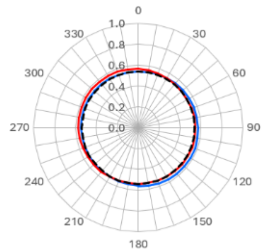
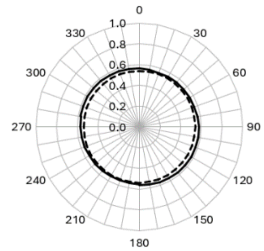
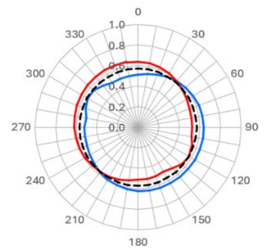
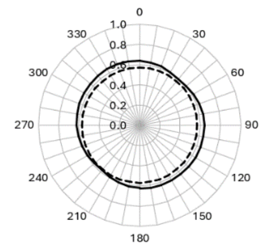
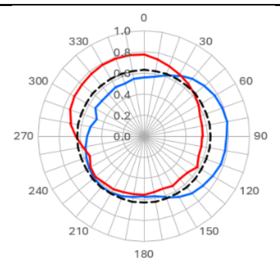
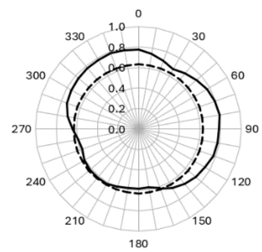
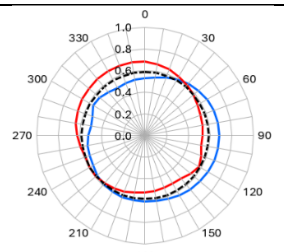
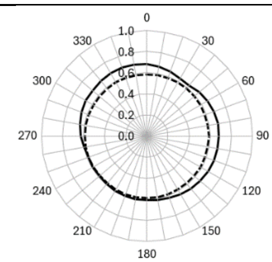
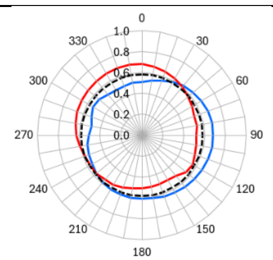
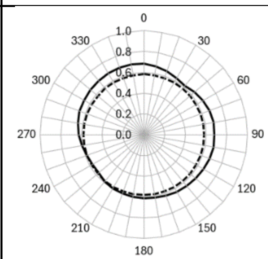
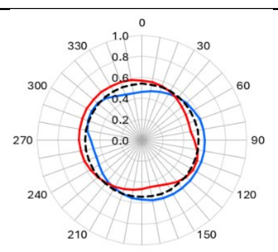
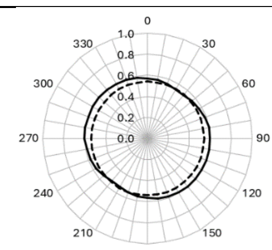
R1 (d=5m)	L,R ear	Best ear	#JND Best ear - Omni
MTI_{low}			Max: 1.33 Avg: 0.92 Min: 0.32
MTI_{mid}			Max: 2.51 Avg: 1.74 Min: 0.20
MTI_{high}			Max: 5.77 Avg: 2.37 Min: -1.80
STI			Max: 3.40 Avg: 1.95 Min: 0.03
STIPA			Max: 3.45 Avg: 2.02 Min: -0.34
D_{50}			Max: 1.33 Avg: 0.76 Min: -0.29

Figure 2: Speech parameter value differences between omnidirectional microphone and HATS for position R1 (sound-receiver-distance: 5 m, viewing angle: 32°)
Left ear Right ear Best ear Omni

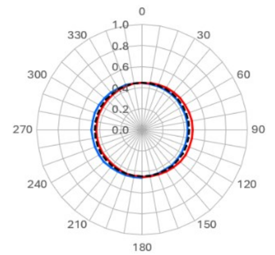
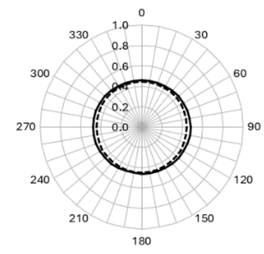
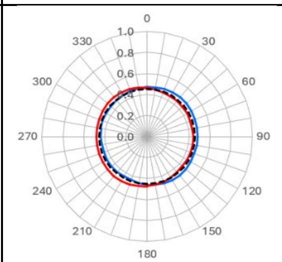
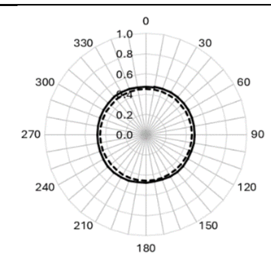
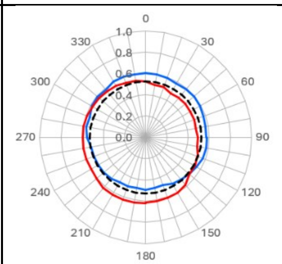
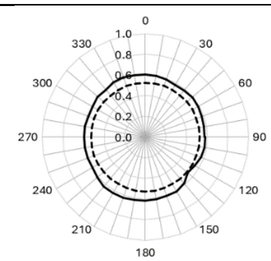
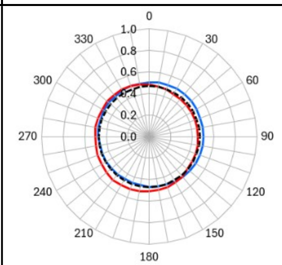
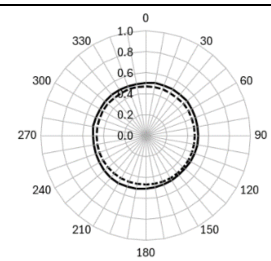
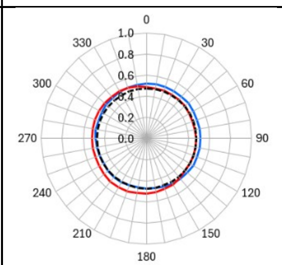
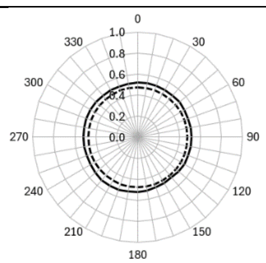
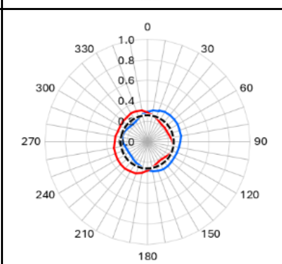
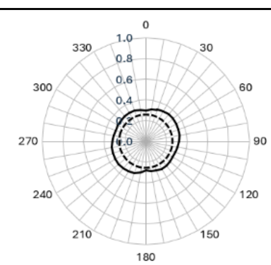
R2 (d=18m)	L,R ear	Best ear	#JND Best ear - Omni
MTI_{low}			Max: 1.19 Avg: 0.76 Min: 0.33
MTI_{mid}			Max: 1.24 Avg: 0.94 Min: 0.47
MTI_{high}			Max: 3.26 Avg: 2.29 Min: 0.40
STI			Max: 1.87 Avg: 1.23 Min: 0.60
STIPA			Max: 1.99 Avg: 1.45 Min: 0.93
D_{50}			Max: 1.70 Avg: 1.23 Min: 0.53

Figure 3: Speech parameter value differences between omnidirectional microphone and HATS for position R2 (sound-receiver-distance: 18 m, viewing angle: 12°)

Left ear — Right ear — Best ear ---- Omni

4. Conclusion and progress

Even for a room designed for an even distribution of the source sound over an audience area and a minimum of produced background noise due to the uniform supply and exhaust of conditioned air, an STI underestimation of more than 3 times the JND (0.1 STI) can occur using the standard monaural (omnidirectional) method as compared to the binaural ('Best Ear') method. For non-reverberant rooms (e.g. open offices) with directional differences in sound intensity for speech and interference sources, the differences in speech intelligibility (or speech privacy) might even be higher. This could be a reason to perform STI measurements in special cases with the Best Ear method using a HATS. Also the use and test of sound masking systems might be more appropriate. To stimulate further research, the Best Ear method will be implemented in DIRAC measurement software. Obtained measurement results can be used for subjective research regarding the relationship between STI(PA) and various intelligibility measures.

REFERENCES

1 IEC 60268-16: 2020: International Standard, Sound System Equipment − *Part 16: Objective rating of speech intelligibility by speech transmission index*. International Organization for Standardization, (2020).

2 Houtgast, T., Steeneken, H.J.M. The modulation transfer function in room acoustics as a predictor of speech intelligibility, *Acustica*, **28** 66-73, (1973).

3 Mapp, P. Is STIPA a robust measure of speech intelligbility performance? AES 118th Convention, Barcelona, Spain, (2005).\

4 Wijngaarden, S., Drullman, R. Binaural intelligibility prediction based on the speech transmission index, *J. Acoust. Soc. Amer.*,**123** 4514-4523, (2008).

5 Hak, C.C.J.M., Haaren, M., Wenmaekers, R.H.C., and van Luxemburg, L.C.J. Measuring room acoustic parameters using a head and torso simulator instead of an omnidirectional microphone, Internoise 2009, Ottawa, Canada, (2009).

6 Hak, C.C.J.M., Wenmaekers, R.H.C., Hak, J.P.M., and van Luxemburg, L.C.J. The source directivity of a dodecahedron sound source determined by stepwise rotation, *Proceedings of Forum Acusticum*, Aalborg, (2011).

7 Schroeder, M.R. Modulation transfer functions: Definition and measurement, *Acustica,* **49** 179-182, (1981).

8 Rife, D.D. Modulation transfer function measurement with maximum-length Sequences, *J. Audio Eng. Soc.*, **40** 779-790, (1992)

9 Hak, C.C.J.M., Wenmaekers, R.H.C. and van Luxemburg, L.C.J. Measuring room impulse responses: Impact of the decay range on derived room acoustic parameters, *Acta Acustica united with Acustica*, **98** 907-915, (2012).

10 ISO 3382-1: International Standard ISO/DIS 3382-1*: Acoustics − Measurement of room acoustic parameters − Part 1: Performance spaces.* International Organization for Standardization, (2009).

11 Bradley, J. S., Reich, R., and Norcrosss, S.G. A just noticeable difference in C_{50} for speech, *Applied Acoustics*, **58** 90-108, (1999).